

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/97748/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Grange, Jacques A. ORCID: <https://orcid.org/0000-0001-5197-249X> and Culling, John F. ORCID: <https://orcid.org/0000-0003-1107-9802> 2016. Head orientation benefit to speech intelligibility in noise for cochlear implant users and in realistic listening conditions. Journal of the Acoustical Society of America 140 (6) , 4061. 10.1121/1.4968515 file

Publishers page: <http://dx.doi.org/10.1121/1.4968515>  
<<http://dx.doi.org/10.1121/1.4968515>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Head orientation benefit to speech intelligibility in noise for cochlear implant users and in realistic listening conditions

Jacques A. Grange<sup>a)</sup> and John F. Culling

*School of Psychology, Cardiff University, 70 Park Place, Cardiff CF103AT, United Kingdom*

(Received 18 May 2016; revised 7 November 2016; accepted 9 November 2016; published online 2 December 2016)

Cochlear implant (CI) users suffer from elevated speech-reception thresholds and may rely on lip reading. Traditional measures of spatial release from masking quantify speech-reception-threshold improvement with azimuthal separation of target speaker and interferers and with the listener facing the target speaker. Substantial benefits of orienting the head away from the target speaker were predicted by a model of spatial release from masking. Audio-only and audio-visual speech-reception thresholds in normal-hearing (NH) listeners and bilateral and unilateral CI users confirmed model predictions of this head-orientation benefit. The benefit ranged 2–5 dB for a modest 30° orientation that did not affect the lip-reading benefit. NH listeners' and CI users' lip-reading benefit measured 3 and 5 dB, respectively. A head-orientation benefit of ~2 dB was also both predicted and observed in NH listeners in realistic simulations of a restaurant listening environment. Exploiting the benefit of head orientation is thus a robust hearing tactic that would benefit both NH listeners and CI users in noisy listening conditions.

© 2016 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[<http://dx.doi.org/10.1121/1.4968515>]

[TCB]

Pages: 4061–4072

## I. INTRODUCTION

Difficulty understanding speech in background noise affects everyone from time to time, but is a particular problem for hearing-impaired listeners. Speech intelligibility is powerfully affected by the speech-to-noise ratio (SNR); just a few decibels can separate perfect comprehension from complete incomprehension. Speech intelligibility in noise can consequently be measured with some precision using a speech reception threshold (SRT), defined as the SNR at which 50% intelligibility is achieved. Hearing impaired listeners often have SRTs only 4–6 dB higher (worse) than normal-hearing (NH) listeners (Plomp, 1986), but this difference is enough to make speech intelligibility in noise their most significant disability (Kramer *et al.*, 1998). Amplification from hearing aids improves speech intelligibility in quiet, but it does not improve SNR and so makes no difference in noise unless the noise is inaudible (Plomp, 1986). Noise reduction algorithms improve SNR. Although they may reduce listening effort (Desjardins and Doherty, 2014), they provide little improvement in intelligibility for listeners with hearing aids, because the speech signal is distorted by the processing (Loizou and Kim, 2011). Cochlear implant (CI) users have even worse problems, with SRTs 10–20 dB higher than NH listeners (Culling *et al.*, 2012). Some noise-reduction algorithms and the use of directional microphones have been shown to provide a benefit for CI users in limited conditions (Hersbach *et al.*, 2012; Mauger

*et al.*, 2012). Any other method of improving SRTs in noise by just a few decibels would provide significant benefits to all listeners, but particularly for users of auditory prostheses.

When speech and noise are spatially separated, there is an improvement in SRT called spatial release from masking (SRM). This effect results from a combination of acoustic differences between the stimulus at each ear and processing of these interaural differences by the brain. It is generally assumed that listeners directly face their conversation partner, and it is thought by both researchers and clinicians that this behavior is most natural (Bronkhorst and Plomp, 1990), most frequently encountered (Koehnke and Besing, 1996), or necessary for lip-reading (Plomp, 1986). However, it would clearly be useful to increase the SRM when possible.

We first noted the potential benefits of head orientation using a computer model of SRM in noise and reverberation (Jelfs *et al.*, 2011; Lavandier and Culling, 2010). The Jelfs *et al.* version of the model is the one used here. The model computes an effective target-to-interferer ratio that is the sum of contributions from two mechanisms. The better-ear path computes the better ear SNR resulting from the head-shadow effect. The binaural-unmasking path computes binaural-masking level differences in each channel from the interaural phase differences between target and masker and from the masker interaural coherence. Both contributions are weighted according to an importance function for speech, before being integrated across frequency bands, then summed. Head orientation affects both contributions to the model by changing target-to-interferer ratio at the ears as well as interaural time delays. The model uses

<sup>a)</sup>Electronic mail: [grangeja@cardiff.ac.uk](mailto:grangeja@cardiff.ac.uk)

binaural-room impulse responses in order to reflect the impact of reverberation, when present. The Jelfs *et al.* model has been validated against a wide variety of SRT data (Culling *et al.*, 2012; Jelfs *et al.*, 2011; Lavandier *et al.*, 2012), predicting the level of SRM in different spatial configurations with different numbers of masking noises and in different levels of reverberation. Increased SRM was predicted when listeners faced a location between the speech source and a single interfering noise source. This prediction is intuitive, because the head acts as an acoustic barrier, and the ear on the side of the speech is shielded from the interfering noise by the acoustic shadow of the head. In addition to this head-shadow effect, the ear on the side of the speech is more sensitive to sound coming from 30° to 60° because the head acts as a baffle and the pinnae increases sensitivity toward the front. Appropriate head orientation to place the speech source in this region of personal space may thus improve speech intelligibility. Existing quantitative studies of head orientation behavior in naturalistic settings have not been analyzed in such a way that they would identify a tendency to orient at 30° away from the target speaker (Ching *et al.*, 2009; Ricketts and Galster, 2008). Most research on SRM assumes that the target speaker will be directly in front of the listener (Beutelmann and Brand, 2006; Bronkhorst and Plomp, 1992; Peissig and Kollmeier, 1997; Plomp, 1986). SRTs are rarely measured with the target speaker in any other location.

The selection of target speech and noise positions can have a substantial impact on the magnitude of SRM. For CI users, SRM is almost always tested speech-facing (i.e., the listener facing the target speaker head on) and with a masker at 90° [see reviews in Van Hoesel (2011) and Culling *et al.* (2012)]. In this configuration and in a sound-treated room, SRM reaches only 3 to 5 dB (e.g., Litovsky *et al.*, 2009). However, three studies have tested CI users in the symmetrical situation where speech and noise sources are placed at equal and opposite azimuths ( $\pm 45^\circ$  or  $\pm 60^\circ$ ) (Culling *et al.*, 2012; Laske *et al.*, 2009; Laszig *et al.*, 2004). These studies demonstrated that with speech and noise sources separated by 90° or 120°, a head orientated midway between the sound sources could lead to a significant head-shadow benefit of bilateral over unilateral implantation (10 to 18 dB). This benefit was defined as the SRT improvement from the spatial configuration that acoustically penalized the better ear (or CI) to the mirror-imaged configuration which favored it. The maximum head-shadow benefit predicted by the Jelfs *et al.* model and experimentally confirmed in Culling *et al.* (2012) is 18 dB for this case.

In a first study focused on the benefit of head orientation to speech intelligibility, Grange and Culling (2016) established a baseline for young NH listeners. In a sound-treated room, we demonstrated that a maximum of 8 dB head-orientation benefit (HOB) was predicted and confirmed to occur at a 60° head orientation when speech and noise were placed at 0° and 180° azimuth, respectively. With the noise placed between 150° and 90°, HOB peaked at 4 to 6 dB at head orientations in the 30° to 45° range. In all these configurations, with noise placed in the rear hemifield, most of the available HOB could be obtained at a 30° head orientation.

None of the studies referred to above tested for audio-visual presentations, despite NH listeners' reliance on lip-reading (Sumby and Pollack, 1954; Summerfield, 1987, 1992) and CI users' higher reliance on lip-reading (Hay-McCutcheon *et al.*, 2005; Rouger *et al.*, 2007; Schorr *et al.*, 2005; Strelnikov *et al.*, 2009) in noisy situations.

The first experiment of the present report aims to show that in situations similar to those described in Grange and Culling (2016), CI users too, can obtain a significant HOB. We also aim to demonstrate that HOB can be obtained at a modest, 30° head orientation that does not detrimentally affect lip-reading, such that head orientation and lip-reading provide cumulative benefits. The second experiment, addresses the potential criticism that such effects are limited to artificial laboratory situations. The effect, while more limited in reverberation, was shown to be robust in real-life situations by creating a very realistic simulation of a restaurant with a target talker sat at the same table as the listener and many other voices distributed around the room.

## II. EXPERIMENT 1

### A. Model predictions and choice of spatial configurations

The choice of spatial configurations was influenced by previous studies of SRM in CI users and informed by predictions from the Jelfs *et al.* (2011) model of SRM.

#### 1. Adequacy of the Jelfs *et al.* model for CI predictions

Culling *et al.* (2012) modified the Jelfs *et al.* model for CI users by removing the binaural unmasking component of the model and obtained a good fit both to their own data and that of Loizou *et al.* (2009). For a bilateral CI user, the model output the better-ear target-to-interferer ratio, assuming equal effectiveness of CIs for speech intelligibility in noise. For a unilateral CI user, the model output the target-to-interferer ratio at their only CI (assuming negligible hearing in the contralateral ear). Here, the Jelfs *et al.* model was used as per Culling *et al.* (2012), with the exception that we used as model input binaural room impulse responses acquired with a head-and-torso simulator in the test environment. Culling *et al.* (2012) argued that the position of a microphone on a processor has a very modest impact on SRM. Incorporating in the model unequal effectiveness of CIs was also found to be unnecessary since it only marginally changed the high correlation between CI data from previous reports and corresponding model predictions. Given the above, no modification of the model was deemed necessary.

#### 2. Selection of spatial configurations

Four spatial configurations were selected: target and masker collocated and in front ( $T_0M_0$ ) served as a reference for SRM data computation; target in front and masker at the rear ( $T_0M_{180}$ ) was predicted to provide the maximum attainable HOB; target in front and masker at the side contralateral to the better ear ( $T_0M_{90}$ ) or on its ipsilateral side ( $T_0M_{270}$ ) were selected because these two configurations were utilized in most prior studies, as discussed in Culling *et al.* (2012).



The three spatially separated configurations are illustrated within each panel of Fig. 1. Jelfs *et al.* model predictions for SRM as a function of head orientation away from the target speaker are shown in the panels of Fig. 1, as derived from binaural room impulse responses acquired in the test environment. These predictions illustrate the benefit of head orientation in each separated spatial configuration for NH listeners and for bilateral (BCI) and unilateral (UCI) CI users, when the left ear (or CI) is the better ear. Arrows highlight SRM for the favorable 30° head orientation at which, according to the model, a large proportion of SRM can be obtained. Where shown, the difference between BCI and NH predictions corresponds to the binaural unmasking contribution to SRM, assumed to be only available to NH listeners; the difference between UCI and BCI predictions corresponds to the predicted benefit of bilateral, over unilateral implantation (see Culling *et al.*, 2012, for in-depth discussion).

In this experiment, the listener either faced the target speaker or faced 30° away (typically favoring the better ear when sources were separated). A modest 30° head orientation was expected to provide a substantial HOB without detrimental impact on the lip-reading. All plots in the results section are transformed to present the left ear as the better ear for speech intelligibility in noise. When the better ear was the right ear, the data were mirrored about the median plane. NH listeners were tested assuming an arbitrary better ear (balanced across participants). Each BCI user's better performing CI in noise was established by comparison of SRTs obtained with speech in front and noise either to the right or to the left in initial practice runs. All CI users were tested in conditions favoring their better or only ear/CI. For UCI users, SRM was additionally measured with the masker at the side ipsilateral to their CI ( $T_0M_{270}$ ). Indeed, even in this worst-case scenario, UCI users were predicted to obtain a large HOB from a modest 30° head turn away from the speech direction.

## B. Materials and methods

### 1. Participants

Ten young NH ( $NH_y$ ) participants, self-reported as normal hearing and aged 18–22 years (mean age 20 years),

were recruited from the Cardiff University undergraduate population (through the School of Psychology's Experimental Management System).

Eight BCI- and nine UCI-user volunteers were recruited from England and Wales through the National CI User Association (NCIUA) and the Cochlear Implant User Group 2004 (Yahoo! CIUG-2004). Table 1 details the specifics of our CI participants. All but one BCI user (B1) had had their last implant fitted at least a year prior to testing and had sequential implantation with the second implant fitted between 2 and 12 years after the first. Participant B1 was simultaneously implanted and had the implants switched on 3 months before testing. All UCI participants had had their implant fitted at least 3 years before testing. All CI users but one (U9) had hardware and software settings such that no microphone directionality was used during testing. Participant U9 used the Esprit 3 G processor from Cochlear. This participant's data will be treated separately as an illustration of the effect of microphone directionality on HOB.

An additional ten NH listeners were recruited from the local Cardiff population, age-matched to the CI users within  $\pm 5$  years. All had normal hearing for their age, as confirmed via pure-tone audiometry screening ( $<20$  dB hearing level from 500 Hz to 4 kHz). From the ten age-matched NH ( $NH_{am}$ ) listeners, a subset was age-matched to each CI user group within 0.5 years on average.

All participants were briefed verbally and in writing prior to signing a consent form. All testing and forms were approved by the Ethics Committee of the Cardiff University School of Psychology.

### 2. Laboratory setup

Two sound-treated rooms were employed, one in Cardiff University (3.2 m  $\times$  4.3 m, 2.6 m ceiling height) and one at University College London (2.7 m  $\times$  4.3 m, 2.2 m ceiling height). Four Minx-10 speakers (Cambridge Audio, London, United Kingdom) fitted 1.3 m above the floor were arranged at cardinal points, at a distance of 1.5 m (Cardiff) and 1.3 m (UCL) from the center of the listener's head. The cross they formed was aligned with the walls and offset to one end of the room such that the rear and side speakers

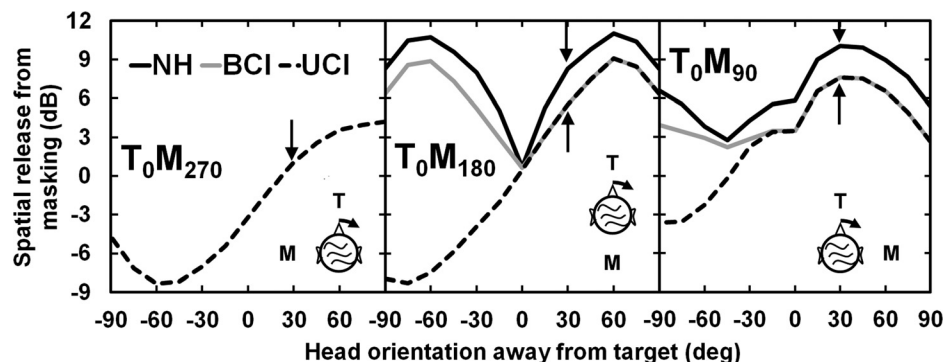


FIG. 1. Jelfs *et al.* (2011) model predictions, from binaural-room-impulse-response acquired in the sound-treated Cardiff room, of spatial release from masking as a function of head orientation away from the target for normal-hearing listeners (NH, solid black line), bilateral (BCI, solid grey line) and unilateral (UCI, dashed black line) CI users at the three separated spatial configurations: target in front and masker at the rear ( $T_0M_{180}$ , center panel), target in front and masker on the side favoring the better ear ( $T_0M_{90}$ , right panel) and target in front and masker on the side ipsilateral to a UCI user's CI ( $T_0M_{270}$ , left panel). All graphs assume the better ear to be the left ear and the arrows point to the prediction for a favorable 30° head orientation.

TABLE I. Specifics of bilateral (B1-8) and unilateral (U1-9) CI-user participants.

CI user	Age	Left CI				Right CI				Aetiology
		Year fitted	Brand	Processor	Implant	Year fitted	Brand	Processor	Implant	
B1	78	2013	Cochlear	Nucleus6	CI-500	2013	Cochlear	Nucleus6	CI-500	Unknown
B2	64	1995	MedEl	Tempo <sup>+</sup>	Pro short-h	2000	MedEl	Tempo <sup>+</sup>	CIS Pro <sup>+</sup>	Meniere
B3	48	2005	Cochlear	Nucleus6	N24	2012	Cochlear	Nucleus6	CI24-RE	Genetic
B4	71	2009	AB	Harmony	HiRes90K	2011	AB	Harmony	HiRes90K	Usher
B5	67	2004	Cochlear	Nucleus5	N24	2006	Cochlear	Nucleus5	CI24-RE	Meniere
B6	66	2001	MedEl	Opus2	Combi40 <sup>+</sup>	2005	MedEl	Opus2	Pulsar	Unknown
B7	66	2001	MedEl	Opus2	Combi40 <sup>+</sup>	2001	MedEl	Opus2	Combi40 <sup>+</sup>	Unknown
B8	78	2007	AB	Harmony	HiRes90K	1995	Cochlear	Freedom	N22	Unknown
U1	39	—	—	—	—	2003	AB	Harmony	C2	Sensorineural
U2	60	2010	MedEl	Opus2	Pulsar	—	—	—	—	Meniere
U3	67	2004	MedEl	Opus2	Combi40 <sup>+</sup>	—	—	—	—	Unknown
U4	67	2008	AB	Harmony	HiRes90K	—	—	—	—	Unknown
U5	32	2004	AB	Harmony	HiRes90K	—	—	—	—	Unknown
U6	74	1996	Cochlear	Nucleus5	N22	—	—	—	—	Streptomycin
U7	59	—	—	—	—	2008	Cochlear	Freedom	N24	Unknown
U8	65	1997	Cochlear	Freedom	N22	—	—	—	—	Unknown
U9	66	2002	Cochlear	Esprit 3 G	N24	—	—	—	—	Viral inf.

were equidistant from the nearest walls and the cross was as remote from the access door as practicable. Each channel of the audio chain was judged to be sufficiently consistent for our purposes in level and spectral response via acquisition of impulse responses and comparison of corresponding excitation patterns (Moore and Glasberg, 1983). The reverberation time (to 60 dB) of both rooms was measured to be approximately 100 ms from the impulse responses, using the reverse integration technique (Schroeder, 1965). The two rooms were acoustically matched as far as practicable with the use of twelve 30 cm × 30 cm foam panels placed where side reflections were most likely to occur. The acoustical matching was judged sufficient for our purpose when the Jelfs *et al.* model predictions in Fig. 1 did not differ by more than 1.2 dB at any point and typically differed by less than 0.5 dB. HOB predictions all differed by less than 0.5 dB. Since all NH listeners and most CI users were tested in the Cardiff room, predictions from binaural room impulse responses obtained in that room were used throughout this report. An adjustable swivel chair was positioned in each room such that regardless of chair rotation, the listener's head was at the center of the loudspeaker array. The experimenter remained in the room at all times, outside of the loudspeaker array and as far as practicable from it. This arrangement was essential to aid interaction with CI users and obtain prompt feedback from them.

The speakers were powered by an Auna six-channel solid-state amplifier (Chal-Tec, Berlin, Germany) driven by a MAYA44USB+ digital-to-analogue converter (ESI AudioTechnik, Leonberg, Germany) connected to a laptop computer. All stimuli were controlled by MATLAB (The MathWorks, Natick, MA) custom-designed programs, making use of the Playrec toolbox (Humphrey 2008–2014); For audio-visual presentations, the speech audio and video streams were synchronized by the VLC program (VideoLAN, Paris, France) and presented on a 17-in. video monitor placed immediately below the 0° azimuth loudspeaker.

### 3. Stimuli

Two SRT protocols were employed, each requiring its own set of stimuli. The first made use of Speech Perception in Noise (SPIN) sentences (Kalikow *et al.*, 1977) recorded audio-visually, so that audio and audio-visual SRTs could be measured and compared. The second employed previously used (Culling *et al.*, 2012; Grange and Culling, 2016) Institute of Electrical and Electronics Engineers (IEEE) sentences from the Harvard corpus (speakers DA and CW) in order to measure more accurate audio SRTs. For the first protocol, a set of 320 high predictability SPIN sentences were audio-visually recorded with an English male speaker (from south-east England). In addition to 200 original SPIN sentences and to complete the set required, 120 new sentences were generated, following the rules established by Kalikow *et al.* (1977). In high-predictability SPIN sentences, the target word is the last word, which is rendered easier to identify by the contextual information that previous words provide. The redundancy of these SPIN sentences was expected to assist CI users and help reduce the standard deviation of SNRs used in the SRT computation. The audio-visual recordings were such that the speaker's face covered two thirds of the video monitor height, delivering a near life-size face. The speaker faced the camera at all times, with his face well lit, for lip-reading purposes. The audio-visual files were batch-processed with FFmpeg (Bellard, 2013) to separate audio and video streams and enable adaptive alteration of sound levels. For the second SRT protocol, a set of 360 IEEE sentences was employed.

All audio files were equalized for root-mean-square power computed over the 3–4 s recordings. The voice associated with each test was utilized to synthesize the masking noise matched in long-term frequency spectrum to that voice. The speech-shaped noise was created using a 512-point finite-impulse-response filter that was based on the calculated excitation pattern of the speech material (Moore and Glasberg, 1983).

#### 4. Audio and audio-visual SRT protocol

Changes were made to our “standard” adaptive threshold method described in Culling *et al.* (2012) in an effort to better adapt the test to CI users. High predictability SPIN sentences (Kalikow *et al.*, 1977) were used instead of IEEE sentences. Initial SNRs were set to  $-18$  dB and  $-4$  dB for NH listeners and CI users, respectively. For the pre-adaptive phase, the SNR increment for each repetition was  $+4$  dB. In the event that the listener failed to recognize the target word after 4 presentations, a new sentence was presented at the previous presentation SNR. The new sentence could be repeated a maximum of 3 times (with  $+4$  dB increments) before being replaced with another sentence (again, with no SNR increment). In fact, none of the listeners required more than two sentences (i.e., more than seven presentations) before recognizing a target word, the trigger required to start the adaptive phase. Once the staircase commenced, SNR was adaptively changed in  $\pm 2$  dB increments, as per the standard protocol. However, each sentence was presented up to three times at increasing SNRs, rather than being renewed at each SNR, until the target word was identified. Repetition of sentences following unsuccessful trials was intended to make more economical use of the relatively small number of audio-visually recorded SPIN sentences. Following Culling *et al.* (2012), the overall sound level throughout an experiment was maintained at 65 dB A (as measured by a digital sound-level meter): an increase in SNR was achieved by simultaneous increase of target level and decrease of masker level, such that overall stimulus level was fixed and could not become uncomfortable. This new protocol is hereafter referred to as the “SPIN<sub>AV</sub> protocol.”

The measurement precision of the SPIN<sub>AV</sub> protocol was compared to that of the standard protocol (that used ten sentences) as a function of the number of sentences used in an audio-only and collocated-source paradigm. The standard deviation of 40 T<sub>0</sub>M<sub>0</sub> SRT measurements per protocol with four NH<sub>y</sub> listeners asymptotized with the SPIN<sub>AV</sub> protocol at the same level (1.9 dB) as the standard protocol when using nine SPIN sentences per run. Nine sentences were therefore used for each SRT-experiment measurement. An SRT offset of  $-1$  dB with the SPIN<sub>AV</sub> protocol compared to the standard was judged inconsequential, given our interest in SRM (i.e., relative) measures. Because of the large number of conditions and to avoid excessively long testing sessions, only two adaptive tracks were performed per condition.

#### 5. Audio-only SRT protocol

Given that only two adaptive tracks per condition in the SPIN<sub>AV</sub> protocol might give rise to substantial data variability, an additional, audio-only protocol was developed that would enable five or six SRT measurements per condition, thereby leading to more accurate SRM measures. The audio-only protocol made use of IEEE sentences, following Grange and Culling (2016), but used the same sentence-substitution regime as the SPIN<sub>AV</sub> protocol. The requirement for triggering the adaptive phase was also relaxed from the recognition of at least two, to the recognition of at least one of the five key words. The remaining sentences in the list of

ten were presented only once following the standard protocol adaptive phase. Here too, the overall sound level was maintained at 65 dB A. This audio-only protocol is hereafter referred to as the “IEEE<sub>A</sub> protocol.”

#### 6. Testing sessions and condition rotation

A first session of SRT measurements employed the SPIN<sub>AV</sub> protocol. The five selected configurations were H<sub>0</sub>M<sub>0</sub>, H<sub>0</sub>M<sub>180</sub>, H<sub>30</sub>M<sub>180</sub>, H<sub>0</sub>M<sub>90</sub>, and H<sub>30</sub>M<sub>90</sub>, where the subscripts denote the head (H) and masker (M) azimuths compared to the target speech. Audio and audio-visual SRTs were measured in separate blocks, each comprised of five spatial configurations. Half of the participants began with an audio-only block, the other half with an audio-visual block, and the sequence of spatial configurations was rotated. The order of the sentence lists remained constant for all participants. Two adaptive tracks were performed and SRTs subsequently averaged between runs.

A second session of SRT measurement in the same five spatial configurations later employed the IEEE<sub>A</sub> protocol. UCI users were also tested in the H<sub>0</sub>M<sub>270</sub> and H<sub>30</sub>M<sub>270</sub> configurations, so that we could explore the potential benefit of head orientation in a spatial configuration that is most detrimental to unilaterally deaf patients. Indeed, placing the masker on the same side as their CI was predicted to lead to negative SRM, if they remained facing the speech. BCI users were also tested in the H<sub>0</sub>M<sub>0</sub>, H<sub>0</sub>M<sub>90</sub>, and H<sub>30</sub>M<sub>180</sub> configurations with each of their implants disabled in turn, which would later enable computation of summation and squelch in these configurations. For NH listeners and UCI users, these configurations were rotated within a block of five and seven configurations, respectively, and the blocks repeated six times. For the BCI users, the monaural conditions were run between binaural blocks and rotated within two dedicated blocks (right, then left CI disabled). All conditions were repeated five times.

#### C. Results

In each (separated) spatial configuration, for each participant and making use of SRTs measured with the IEEE<sub>A</sub> protocol, (1) speech-facing SRM was computed as the speech-facing SRT (condition H<sub>0</sub>M <sub>$\alpha \neq 0$</sub> ) subtracted from the collocated SRT (condition H<sub>0</sub>M<sub>0</sub>) and (2) HOB was computed as the 30° head-orientation SRT (condition H<sub>30</sub>M <sub>$\alpha \neq 0$</sub> ) subtracted from the speech-facing SRT (condition H<sub>0</sub>M <sub>$\alpha \neq 0$</sub> ). Consequently, the sum of speech-facing SRM and HOB is the SRM resulting from concurrent spatial separation of sound sources and 30° head orientation. As such, speech-facing SRM and HOB can be displayed as cumulative measures. Figure 2 displays speech-facing SRM (lower panels), HOB (middle panels) and their cumulative effect (upper panels) averaged within each listener group for all three separated spatial configurations. The standard error of group means did not exceed 1 dB and averaged 0.65, 0.38, 0.55, and 0.63 dB for NH<sub>y</sub> and NH<sub>am</sub> listeners and BCI and UCI users, respectively. The isolated directional microphone case (UCId) had a mean standard error of 1 dB (across five repeat runs). SRM and HOB outcomes are compared below to Jelfs *et al.* (2011)



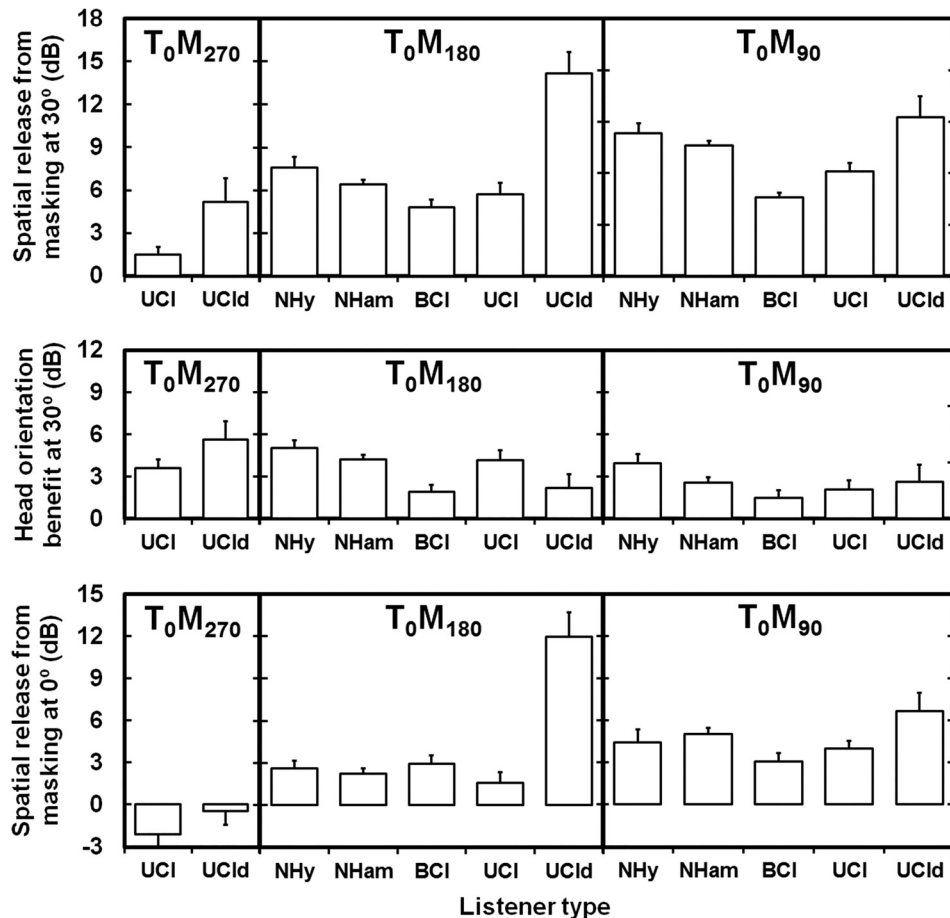


FIG. 2. Speech-facing SRM (bottom panels), head-orientation benefit (middle panels) from a beneficial 30° head orientation away from the speech and SRM resulting from the combination of source separation with a 30° head orientation away from the speech, as measured in each of the three separated spatial configurations [ $T_0M_{270}$  (left panels),  $T_0M_{180}$  (center panels) and  $T_0M_{90}$  (right panels)] and for each listener group [young NH adults ( $NH_y$ ); bilateral and unilateral CI users (BCI and UCI); a single unilateral CI user with directional microphone enabled ( $UCId$ ); NH adults age-matched to the CI users ( $NH_{am}$ )]. Speech-facing SRM is the benefit of spatial separation of target and masker, when the listener faces the target speaker. HOB is the additional benefit of a 30° head orientation with the same spatial separation. Consequently, the sum of speech-facing SRM and HOB is the SRM resulting from concurrent spatial separation and head orientation. Error bars denote standard error of cross-participant means, except for the unilateral CI user with a directional microphone, where error bars denote standard error of within-participant means.

model predictions computed from binaural-room impulse responses acquired in the Cardiff test room. Any concern relating to young NH listeners not having been specifically screened for hearing loss was alleviated by the standard deviation of audio-only SRTs averaged across spatial configurations being as low as 0.6 dB (1.7 dB range).

### 1. Speech-facing SRM

At  $T_0M_{180}$  and for all groups, speech-facing SRM was large (1.6–2.6 dB) compared to the 0.5–0.7 dB predicted by the model. At  $T_0M_{90}$ , speech-facing SRM measured 3.1–5.1 dB and compared favorably with predictions for all groups (within 0.4–1.4 dB). Speech-facing SRM was increased by 1.5–10 dB with a directional microphone, depending on masker location. At  $T_0M_{270}$ , UCI users' speech-facing SRM measured –2.1 dB and was comparable to prediction (–3.2 dB). Analyses of variance (ANOVAs) operated within each listener group on speech-facing SRTs confirmed a significant effect of masker separation [ $NH_y$   $F(2,18) = 15.8$ ;  $NH_{am}$   $F(2,18) = 77.0$ ; BCI  $F(2,14) = 17.7$ ; UCI  $F(3,21) = 30.4$ ,  $p \leq 0.001$  for all groups] with pairwise comparisons between collocated and separated conditions within each group all showing a significant difference, i.e., speech-facing SRM ( $p < 0.02$  for all pairs, based on estimated marginal means with no adjustments), except for unilateral CI users at  $T_0M_{180}$ . Speech-facing SRM obtained by  $NH_y$  and  $NH_{am}$  adults was compared with an ANOVA and not found to differ [ $F(1,18) = 0.35$ ,  $p > 0.5$ ].

### 2. Head-orientation benefit

At  $T_0M_{180}$ , HOB measured 1.9 to 5.0 dB across groups and was notably smaller than predicted by the model (5.0 to 7.6 dB). At  $T_0M_{90}$ , HOB measured 1.5 to 3.9 dB and was comparable to the prediction (4.1 dB), except for BCI users. Overall, BCI users obtained notably less HOB than predicted. At  $T_0M_{270}$ , UCI users' HOB measured 3.6 dB and was comparable to the prediction (4.3 dB). Across listener groups and configurations, 30° HOB was confirmed significant by an ANOVA that compared SRM between head orientations [ $F(1,32) = 338.2$ ,  $p < 0.001$ ]. HOB was confirmed significant within each listener group by separate ANOVAs [ $NH_y$   $F(1,9) = 146.4$ ;  $NH_{am}$   $F(1,9) = 141.0$ ; BCI  $F(1,7) = 18.9$ ; UCI  $F(1,7) = 129.2$ ,  $p \leq 0.005$  for all groups].

### 3. Cumulative effect of masker separation and 30° head orientation on SRM

For NH listeners, adding speech-facing SRM and HOB led to SRM in reasonably good agreement with model predictions at  $T_0M_{180}$  (6.4 and 7.6 dB for  $NH_y$  and  $NH_{am}$  listeners, respectively, versus 8.3 dB predicted) and at  $T_0M_{90}$  (7.6 and 8.4 dB for  $NH_y$  and  $NH_{am}$  listeners, respectively, versus 10 dB predicted), but older NH adults obtained less SRM than their younger counterparts in both conditions. For UCI users, cumulative SRM was again in good agreement with predictions (1.5, 5.6, and 6.1 dB versus predicted 1.1, 5.5, and 7.6 dB at  $T_0M_{270}$ ,  $T_0M_{180}$ , and  $T_0M_{90}$ , respectively). For BCI users, cumulative SRM was lower than predicted (4.8

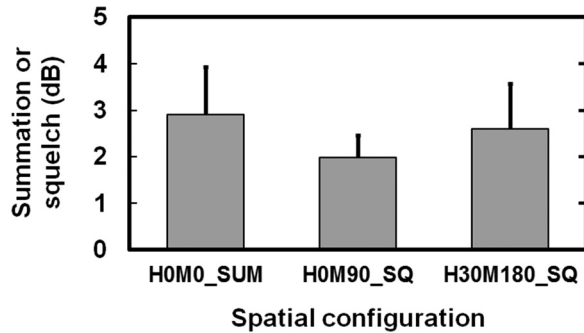


FIG. 3. Measures of summation in the collocated configuration ( $H_0M_0$ \_SUM label) and squelch in separated configurations ( $H_0M_0$ \_SQ and  $H_{30}M_{180}$ \_SQ labels), averaged across bilateral CI users and defined as the benefit of activating the poorer CI in addition to the better CI (the CI that provides the better speech-in-noise intelligibility). Error bars are standard errors of the means.

and 4.6 dB versus 5.5. and 7.6 dB at  $T_0M_{180}$  and  $T_0M_{90}$ , respectively), primarily due to their HOB being lower than other listeners’.

#### 4. The directional microphone case

As can be seen in Fig. 2, speech-facing SRM increased by 10 dB at  $T_0M_{180}$  in our directional-microphone UCI user case, compared to the omnidirectional-microphone UCI user group mean. At  $T_0M_{90}$ , speech-facing SRM was also increased by nearly 1.5 dB. A significant HOB was found in all configurations; although it was reduced a little compared to that of omnidirectional UCI users.

#### 5. BCI users’ summation and squelch

Summation is defined here as the  $H_0M_0$  SRT improvement found when activating the worse-performing CI in addition to activating only the best-performing CI. Squelch is defined as the same benefit, but for spatially separated sound sources. Squelch is traditionally measured in the  $H_0M_{90}$  configuration, where only the masker signal is subject to interaural level differences. We measured it also in the  $H_{30}M_{180}$  configuration, where both speech and noise signals differ between ears. Summation and squelch outcomes are plotted in Fig. 3, as extracted from SRTs acquired with the IEEE<sub>A</sub> protocol. An average summation of 2.9 dB (1 dB standard error) was measured while squelch was 2.0 and 2.6 dB (0.5 and 1 dB standard error) at  $H_0M_{90}$  and  $H_{30}M_{180}$ ,

respectively. A within-subject t-test (2-tailed) comparing  $H_0M_0$  SRTs with both CIs enabled to SRTs with the best CI enabled showed the summation effect to be significant [ $t(7) = 2.84$ ,  $p < 0.025$ ]. The squelch effect was also significant at  $H_0M_{90}$  [ $t(7) = 4.05$ ,  $p < 0.01$ ] and at  $H_{30}M_{180}$  [ $t(7) = 2.68$ ,  $p < 0.05$ ].

#### 6. Lip-reading benefit

In each spatial configuration, for each participant and making use of SRTs measured with the SPIN<sub>AV</sub> protocol, the lip-reading benefit was computed as the audio-visual SRT subtracted from the audio-only SRT. Figure 4 displays lip-reading averaged within each listener group for the five configurations common to all groups ( $H_0M_0$ ,  $H_0M_{180}$ ,  $H_{30}M_{180}$ ,  $H_0M_{90}$ , and  $H_{30}M_{90}$ ). The benefit of lip-reading measured typically 3 dB for NH listeners and 5 dB for CI users. Across listener groups and spatial configurations, an ANOVA for SRTs in the two presentation modalities confirmed a significant benefit of visual cues [ $F(1,32) = 368.9$ ,  $p < 0.001$ ]. An interaction between modality (audio or audio-visual) and listener type indicates that CI users are better lip-readers and/or more dependent on visual cues [ $F(3,32) = 7.45$ ,  $p < 0.001$ ]. The lack of interaction between modality and spatial configuration [ $F(4,128) = 0.56$ ,  $p = 0.69$ ] indicated that configuration had no impact on lip-reading. Most relevant to our study was that a 30° head turn had no detrimental effect on lip-reading within each group [ $NH_y$   $F(1,9) = 0.77$ ,  $p = 0.40$ ;  $NH_{am}$   $F(1,9) = 0.18$ ,  $p = 0.68$ ; BCI  $F(1,7) = 0.26$ ,  $p = 0.62$ ; UCI  $F(1,7) = 0.23$ ,  $p = 0.65$ ]. Thus, a sidelong regard, i.e., orienting the gaze to compensate for a modest head orientation away from the target speaker, facilitates a significant benefit of head orientation, additive to that of lip-reading.

### III. EXPERIMENT 2

Experiment 1 demonstrated the effectiveness of head orientation in a sound-treated room with a single interfering sound source. It also showed that the benefit of lip-reading is robust to head rotation of at least 30°. In a real listening environment, such as a bar or restaurant, there are likely to be multiple interfering sounds sources and there will certainly be reverberation. The second experiment addresses the question of whether the head-orientation benefit still occurs in such an environment. The approach taken is to simulate, as

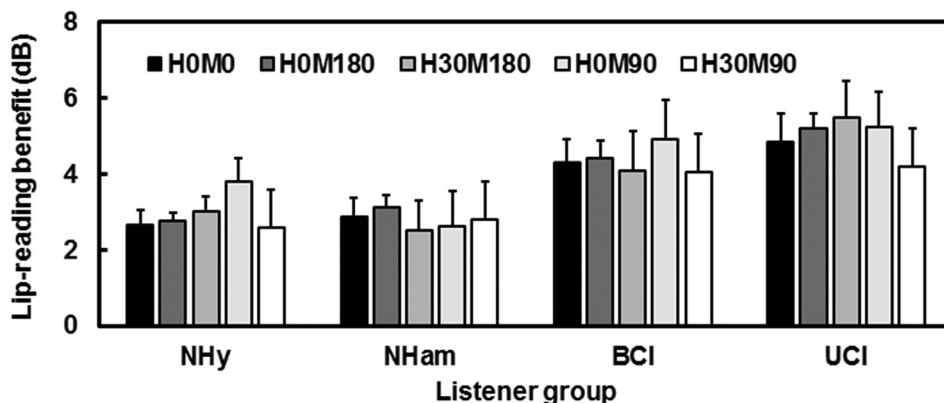


FIG. 4. Lip-reading benefit computed as threshold improvement from audio to audio-visual conditions and averaged within each listener group in five spatial configurations ( $H_0M_0$ ,  $H_0M_{180}$ ,  $H_{30}M_{180}$ ,  $H_0M_{90}$  and  $H_{30}M_{90}$ ). Error bars are standard errors of the means.



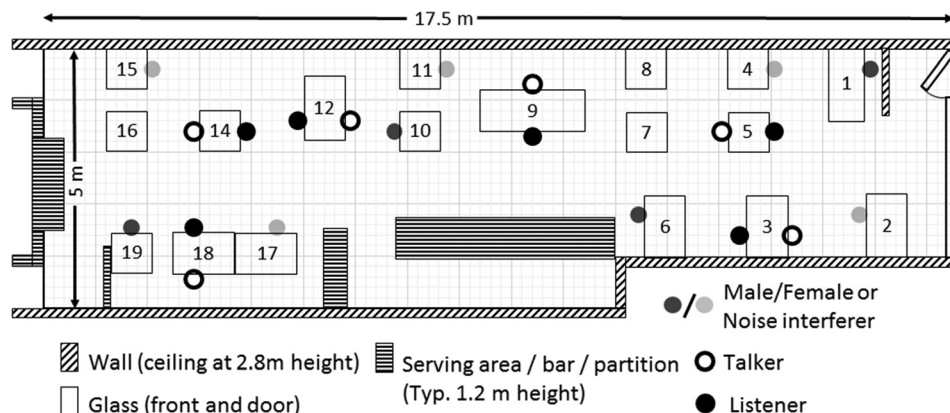


FIG. 5. Plan view of the Mezzaluna restaurant (Cardiff) where impulse responses were acquired from 18 different listener seats and with 18 talker or interferer (opposite) seats. Black-filled circles highlight the listener positions tested for, light-grey-filled circles the noise or female-voice interferer, dark-grey-filled circles the additional noise or male-voice interferers and the open circles the target male talkers facing listener positions.

realistically as possible, a restaurant listening situation using a methodology similar to that of Culling (2016). A virtual simulation was created of a real restaurant, and the effect of head orientation in this virtual environment was measured.

## A. Materials and methods

### 1. Participants

Sixteen young, self-reported NH adults, aged 18–21 years (mean age 20.2 years) were recruited in the same manner as NH<sub>y</sub> participants of experiment 1 and participated in a 90-min session.

### 2. Stimuli and methods

The virtual simulated restaurant was created by convolving dry speech (i.e., without reverberation) with binaural-room impulse responses. The 475-ms impulse responses were recorded in a Cardiff restaurant (Fig. 5) during its closing hours using the tone-sweep method (Farina, 2007; Müller and Massarani, 2001). Ten-second exponential tone sweeps were presented from a Minx-10 loudspeaker (Cambridge Audio, London, United Kingdom) to a B&K-4100 head and torso simulator (Brüel & Kjær, Nærum, Denmark). Source and receiver locations were chosen directly opposite each other at each of 18 tables in the restaurant. Impulse responses were recorded between every combination of source and receiver locations. The head of the B&K simulator was also oriented to each of three positions ( $-30^\circ$ ,  $0^\circ$ ,  $30^\circ$ ). Thus, a total of 18 source positions  $\times$  18 receiver positions  $\times$  3 head orientations = 972 impulse responses were recorded. A subset of 180 impulse responses were needed in this experiment.

In the simulations, the listener was seated at one of six tables and adopted each of the three head orientations at each table. Target speech was presented from the seat opposite. Nine interfering voices (five female and four male) with British accents, or nine interfering speech-shaped noises were distributed in a randomly selected, but fixed configuration across other tables (see Fig. 5). SRTs were measured with stimuli presented over headphones and using Harvard IEEE sentences standard methods (Culling and Mansell, 2013; Plomp and Mimpen, 1979) except that the interfering sources produced continuous speech or noise. Ten sentences were used to obtain an SRT. The interfering speech was

taken from book readings posted on [librivox.org](http://librivox.org). The interfering noises were filtered to match the interfering voices in excitation pattern.

SRTs were measured for 6 listener positions  $\times$  3 head orientations  $\times$  2 interferer types = 36 conditions with 36 lists of ten sentences. Listeners were familiarized with the procedure by two practice runs with a single interfering noise, using spatial configurations different from those used in the experiment. Because of the large number of conditions, each participant received a random sequence of conditions, while the sentences were presented in a fixed order.

## B. Results

Figure 6 shows the mean SRTs for each table, head orientation and interferer type (symbols). Also shown are predictions based on the Jelfs *et al.* (2011) model of speech reception in noise and reverberation (lines). It can be seen that SRTs are highest when the listener directly faces the speech source in the majority of cases. An analysis of variance for SRT, with factors listener table number, head

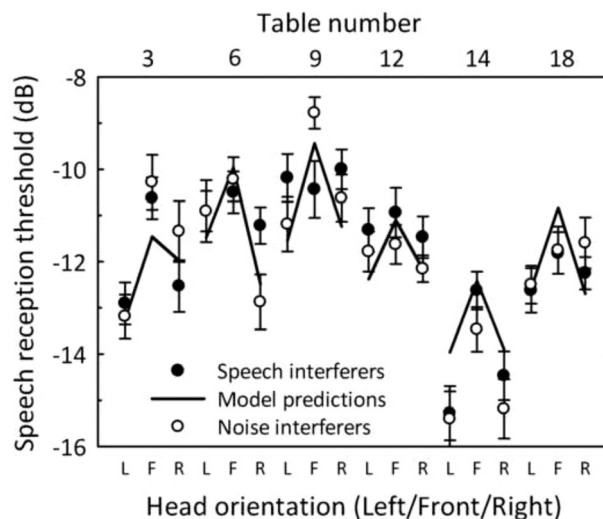


FIG. 6. SRTs obtained in situations with left ( $-30^\circ$ )/front ( $0^\circ$ )/right ( $+30^\circ$ ) head orientations (L/R labels, on the lower horizontal axis) for each of the listener/talker pairs (at Tables 3, 6, 9, 12, 14, and 18, labels on the upper horizontal axis) and with speech (black-filled circles) or noise (open circles) interferers. Error bars are standard errors of the means. Black lines represent model predictions with their mean equalized to that of the noise-masker conditions.

orientation, and interferer type, confirmed a significant benefit of head orientation [ $F(2,30) = 23.3$ ,  $p < 0.001$ ]. From Fig. 6, orienting  $30^\circ$  away from the target source improved speech reception in speech-shaped noise (open symbols) in each listening position, in line with the predictions of the Jelfs *et al.* model. When interfering speech was used (filled symbols), the picture was a little more mixed, but shows the same average pattern, and the interaction between head orientation and interferer type was not significant. SRTs in speech and noise did not differ significantly. A main effect of table number [ $F(5,75) = 53.7$ ,  $p < 0.001$ ] revealed that there are systematic differences between listening positions with some seats in the restaurant allowing lower SRTs than others. Averaging the mean SRTs for speech and noise, a strong correlation between data and predictions [ $r(1,17) = 0.88$ ,  $p < 0.001$ ] confirmed that the model also predicts the variations across tables and head orientations accurately.

## IV. DISCUSSION

SRTs measured in a sound-treated environment confirmed the predicted benefit to speech intelligibility in noise of a modest ( $30^\circ$ ) head orientation away from a talker when a single steady-noise interferer is azimuthally separated from the speech by  $180^\circ$  or  $90^\circ$ . This HOB was significant for normal-hearing listeners (3–5 dB) as well as for UCI users (2.5–5 dB) and BCI users (1.5–2.5 dB). The lip-reading benefit extracted from comparing audio-visual to audio-only outcomes was significant and somewhat larger in CI users (5 dB) than in NH listeners (3 dB). Crucially, lip-reading was not detrimentally affected by a  $30^\circ$  head orientation. The SRT data therefore showed that significant HOB can be exploited by CI users, in addition to the lip-reading that non-blind hearing-impaired listeners rely on. Data from a UCI user that made use of a directional microphone suggest that a directional microphone does not remove this HOB.

### A. Speech-facing SRM and HOB

The speech-facing SRMs for NHy listeners (2.6 dB at  $T_0M_{180}$  and 4.4 dB at  $T_0M_{90}$ ) were in reasonable agreement with those obtained by Plomp (1976), 3.0 and 5.4 dB, respectively. SRM obtained with our CI participants at the typical  $H_0M_{90}$  configuration (3–4 dB) falls within the range covered by previous reports and reviewed in Culling *et al.* (2012), although BCI users' SRM is on the low end. The head-shadow effect measured from our UCI users (6 dB) also falls in the range covered by previous reports and reviewed by Van Hoesel (2011) and is a very good match to that measured by Culling *et al.* (2012). Summation and squelch results are compared with the results from Litovsky *et al.* (2006) in the bilateral-CI-users section below.

#### 1. Addressing the main discrepancy with model predictions

The  $T_0M_{180}$  speech-facing SRM was higher across all listener groups than predicted by the model. Since the prediction was based on acoustic measurements of the sound-

treated room itself, the result cannot be explained by modest reverberation in that room. When facing the speech, there is a sharp predicted improvement in SRT for any deviation in correct head orientation. As a result, the measured SRTs should be reduced by any misalignment of the head. In contrast, for other head orientations the predicted SRT changes in different directions with head misalignment, so the SRT measurements are not biased by random misalignments. Misalignment of the head orientation during the SRT runs thus seems the most likely explanation for the high speech-facing SRM in  $T_0M_{180}$  (see also Grange and Culling, 2016). The fact that UCI users (the only listeners predicted not to gain HOB by turning either way, see Fig. 1) obtained by far the lowest  $T_0M_{180}$  speech-facing SRM (see Fig. 2) reinforces the above interpretation of the data.

## 2. Group differences

The measures of SRM in configurations that facilitate binaural unmasking were lower for CI users than for NH listeners, which is consistent with the assumption made that CI users do not benefit from binaural unmasking. Both CI users and NHy also had lower HOB than predicted. If, as argued above, the  $T_0M_{180}$  speech-facing SRM was inflated by head misalignment, 1–2 dB of the measured  $T_0M_{180}$  speech-facing SRM may in fact have been HOB. This misattribution would account for a deflated measure of  $T_0M_{180}$  HOB. However, it does not fully account for the reduced HOB in  $NH_{am}$  listeners. These older, NH adults may have suffered from a loss of binaural unmasking, consistent with recent reports of an age-related decline in the binaural processing of temporal envelope and fine structure (King *et al.*, 2014; Moore *et al.*, 2012; Hopkins and Moore, 2011) that reduced their HOB and their overall SRM.

The case of the UCI user who used a directional microphone setting demonstrated how, by suppressing sound waves coming from the rear, the  $T_0M_{180}$  speech-facing SRM was increased by over 10 dB for  $T_0M_{180}$ . However, the  $T_0M_{90}$  and  $T_0M_{270}$  speech-facing SRM values were increased by only 1.5 dB. Thus, if the masker were placed in the frontal hemifield, SRM was hardly affected by the sensitivity pattern of a directional microphone. Just as importantly, a significant  $30^\circ$  HOB remained in all three configurations, so microphone directionality does not remove HOB. This result is also predicted by the model, because the diffracting effects of the head alter the directional microphone sensitivity pattern to favor sounds  $30^\circ$ – $40^\circ$  away from the front. Figure 7 illustrates the effect of the head with the speech-weighted directional response of *in situ* directional microphones. These predictions were based on measurements of head-related impulse responses from the microphones of Oticon behind-the-ear hearing aids, placed on an acoustic manikin. The directional patterns in Fig. 7 represent only an illustrative example rather than the particular fixed directional pattern that would be produced by the Esprit 3 G processor, or the directional pattern that would be produced by the Oticon hearing aid on which it is based. Nonetheless, they capture an asymmetry in the left- and right-ear responses that would be common to any

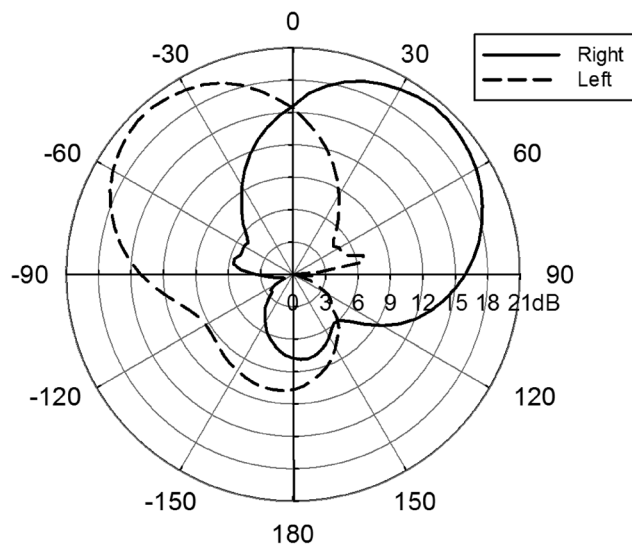


FIG. 7. Sensitivity patterns of *in situ* directional microphones, generated from a simple broadband delay-and-subtract operation on impulse responses acquired from the two microphones of an Oticon behind-the-ear hearing aid fitted either side of an acoustic manikin. This figure aims to illustrate that a directional pattern is modified by the head-shadow in such a way that sensitivity maxima sit in the  $\pm 30^\circ$ – $50^\circ$  regions.

two-port *in situ* directional microphone which produces a stronger response to sounds from  $\pm 30^\circ$ – $50^\circ$ . It should be noted that this “distortion” in the directional pattern is probably a desirable feature for bilaterally implanted patients, because it reflects the fact that interaural level differences are preserved.

### 3. Bilateral CI users

BCI users stood out in that their measured HOB was less than half of model predictions. At  $T_0M_{180}$  this outcome may again be explained by inaccuracies in head orientations during testing. However, at  $T_0M_{90}$ , the HOB shortfall clearly requires another explanation, because the overall SRM sits 3 dB lower than predicted. Additional measures of summation (2.9 dB at  $H_0M_0$ ) and squelch (2.0 dB at  $H_0M_{90}$  and 2.6 dB at  $H_{30}M_{180}$ ) from BCI users were found to be significantly larger than previously reported in the literature. These correspond to the “diotic” and “binaural” benefits reviewed by Van Hoesel (2011). Compared to summation outcomes reported in the Litovsky *et al.* (2006) multi-center study (the effect they call *binaural redundancy*), our mean summation seems larger than their 1.5 dB, but their range,  $-6$  to  $+9$  dB, was comparable to ours,  $-3.5$  to  $+6.5$  dB. Given their much larger sample, and standard errors being large (1 dB) in both studies, the difference is probably not significant. Their measure of squelch matched ours, at 2 dB. Consistently with Litovsky *et al.* (2006), binaural summation or squelch effect size in BCI users was much smaller than the  $T_0M_{90}$  SRM of our BCI users or the  $T_0M_{90}$  head-shadow effect of our UCI users.

Assuming BCI users do not benefit from binaural unmasking, both summation and squelch are believed here to be due to the information provided by the two CIs differing in spectral content, in a complementary manner such that

spectral summation occurs. Our middle-aged or older BCI users are unlikely to have equal nerve survival along their spiral ganglia, and some CI electrodes may be disabled, so as to prevent, for instance, unintended facial nerve excitation. It is therefore plausible that their two CIs deliver information from complementary spectral regions. The model ignores the SNR at the poorer ear, but the poorer ear could still be relevant to speech intelligibility if it contains such complementary spectral information (Culling *et al.*, 2012).

HOB may have been lower in BCI than in UCI users because BCI users already benefit from spectral summation when facing the speech and turning away from the speech might reduce the summation effect. Indeed, spectral summation should be maximum when SNRs at the two ears are similar. Orienting the head so as to bring the better ear closer to the target speech will not only improve the SNR at the better ear as the model predicts, it will also reduce the SNR at the poorer ear, thereby reducing the benefit of providing the speech information from that ear to the brain. Even if summation occurred only as a result of a reduction of internal noise at a central auditory brain level, the same principle would apply. The fact that with an additional CI, BCI users’ SRM obtained with a  $30^\circ$  head turn is lower than UCI users’ in both spatial configurations (by up to 1.5 dB at  $H_{30}M_{90}$ ) further reinforces the above interpretation of the data. It therefore seems that BCI users’ HOB can be reduced by a loss of summation in some spatial configurations.

### B. Reliance on lip-reading

A sidelong regard with a head orientation of  $30^\circ$  maintained the benefit of lip-reading at the same level as when directly facing the speaker. A linear regression analysis of lip-reading benefit versus  $H_0M_0$  audio-only SRTs showed a negative correlation between proficiency of listeners in recognizing speech in noise and the added benefit of visual cues ( $r = 0.66$ ,  $t = 4.31$ ,  $p < 0.001$ ). This correlation is not surprising since an elevation in listeners’ audio-only SRT will increase their reliance on lip reading and also can motivate individuals to improve their lip-reading skills (e.g., Strelnikov *et al.*, 2009). Every 6 dB in SRT elevation was partially compensated for by 1 dB improvement in lip-reading benefit. Since talkers differ in the ease with which they can be lip-read, the regression slope of data acquired with a different talker could be significantly different to the slope we found. One might expect that the easier the talker is to lip-read, the higher the slope. Thus, for more familiar talkers, lip-reading might go much further toward compensating for the threshold elevation CI users suffer from. Previous studies also showed that the lip-reading benefit is highly dependent on the ease of lip-reading of the sentence material (Macleod and Summerfield, 1987). To date, it has not been established whether stimulus material and talker contributions to the ease of lip-reading are independent or interact.

### C. Realistic listening conditions

Experiment 2 examined HOB in realistic listening conditions, and showed that consistent benefits exist in the presence of multiple interferers and reverberation. One might



imagine that the effect of such distributed interference would be to suppress any effects based on head-shadow and better-ear listening, because both ears would receive roughly the same level of noise. Indeed, Hawley *et al.* (2004) and Culling *et al.* (2004) showed that if just two or three nearby interfering sources are located in different hemifields, effects attributable to better-ear listening become negligible. However, SNR depends on the levels of both the speech and the noise. While many of the interfering sound sources in a noisy room are in the reverberant field and consequently reach both ears at a similar sound level, the target speech is usually close by, in the direct field, and reaches the nearer ear at a higher sound level. Here, the benefit of “head-shadow” is not a shadowing effect at all, but the amplification of a target wave of near-normal incidence reflecting back on itself after bouncing off the surface of the head. By turning the head, one can place one ear into this amplified part of the target’s sound field. This benefit should occur for practically any listening situation and practically any listener, provided the target source is close.

The reader might consider the sidelong-regard posture unnatural or more effortful. Informal feedback from all CI users who participated in the study was that they did not perceive this strategy to be an issue for them or for familiar conversation partners. They actually welcomed it. In addition, it is not uncommon for listeners to instinctively use a sidelong regard in noisy situations. This strategy is common place in loud industrial settings, for instance. The human oculomotor range is limited to a  $\pm 55^\circ$  eye-in-head lateral angle (Guittun and Volle, 1987). Although maintaining a lateral angle up to  $30^\circ$  may be more effortful than viewing the speaker’s face head-on, we feel that HOB will outweigh the potential extra effort. This expectation remains to be confirmed.

#### D. Importance of our findings to the hearing impaired

CI users are known to struggle to understand speech in noisy social settings. Despite all the recent efforts made to restore access to interaural time delays at low frequencies, BCI users exhibit negligible binaural unmasking and pitch cues are limited by the relatively sparse encoding of sound by CIs. As a result, CI users only benefit from head-shadow and lip-reading benefit effects, binaural unmasking being inaccessible (Churchill *et al.*, 2014; Van Hoesel *et al.*, 2008) and discrimination of voice fundamental frequencies very limited (Carroll and Zeng, 2007; Geurts and Wouters, 2004). Dip-listening is also much harder for CI users (Nelson *et al.*, 2003). Given the limited cues available to CI users, any guidance about how to optimally combine head-orientation and lip-reading benefits could be highly valuable to them. Such guidance could make the difference between social isolation and active enjoyment of social interactions. As guidance may benefit interactions with a familiar, easier-to-lip-read conversation partner, it is even more critically important for unfamiliar, harder-to-lip-read conversation partners. While the research presented here focusses on CI users, it can equally well serve to help other hearing-impaired listeners, whether partially and/or unilaterally deaf. Since binaural unmasking represents a small part of a NH

listener’s SRM and hearing-impaired listeners often exhibit a reduction in binaural unmasking, the conclusions drawn from the present studies may transfer to hearing aid users as well as unaided hearing-impaired listeners.

#### V. CONCLUSION

The presented study has shown that there is a substantial head-orientation benefit available to CI users’ speech understanding in noise. In sound-treated rooms, NH listeners obtained a large benefit, which was somewhat reduced by a loss of binaural unmasking in the older NH adults, who were age-matched to our CI user participants. Despite the absence of binaural unmasking in unilateral CI users, their head-orientation benefit matched that of young NH listeners (5 dB) with the masker initially at the rear. The benefit was reduced, but still significant with the masker initially to the side contralateral to their CI (2.5 dB). Bilateral CI users exhibited the lowest benefit of head orientation, presumably because they already benefitted from substantial spectral summation. A modest  $30^\circ$  head orientation did not affect the lip-reading benefit measured in NH listeners (3 dB) and CI users (5 dB). Head orientation up to  $30^\circ$  and lip-reading therefore provide cumulative benefits. In normal-hearing listeners, head-orientation benefit of  $>1$  dB was found to be robust in a realistic listening environment with multiple interfering sounds sources (speech-shaped noises or voices) and reverberation. These findings with CI users and NH listeners may extend to other hearing-impaired listeners, so all listeners can enjoy the benefits of the sidelong regard in noisy environments.

#### ACKNOWLEDGMENTS

We would like to thank our participants for their time, Action on Hearing Loss for funding part of this research, the management and staff at the Department of Speech, Hearing and Phonetic Sciences of University College London for providing access to a test room in their Chandler House laboratories, the owners of Mezzaluna for kindly giving us access to their restaurant and Lilith Ramage, Amy Shields, Conor Czech, and Emer Hammond for their assistance in gathering acoustic data.

- Bellard, F. (2013). “FFmpeg,” [www.ffmpeg.org](http://www.ffmpeg.org) (Last viewed June 2, 2013).
- Beutelmann, R., and Brand, T. (2006). “Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **120**, 331–342.
- Bronkhorst, A., and Plomp, R. (1990). “A clinical test for the assessment of binaural speech perception in noise,” *Int. J. Audiol.* **29**, 275–285.
- Bronkhorst, A., and Plomp, R. (1992). “Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing,” *J. Acoust. Soc. Am.* **92**, 3132–3139.
- Carroll, J., and Zeng, F. (2007). “Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations,” *Hear. Res.* **231**, 42–53.
- Ching, T., Brien, A., Dillon, H., and Hain, J. (2009). “Directional effects on infants and young children in real life: Implications for amplification,” *J. Speech Lang. Hear. Res.* **52**, 1241–1254.
- Churchill, T., Kan, A., Goupell, M., Ihlefeld, A., and Litovsky, R. (2014). “Speech perception in noise with a harmonic complex excited vocoder,” *J. Assoc. Res. Otolaryngol.* **15**, 265–278.

- Culling, J., Hawley, M., and Litovsky, R. (2004). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," *J. Acoust. Soc. Am.* **116**, 1057–1065.
- Culling, J., Jelfs, S., Talbert, A., Grange, J., and Backhouse, S. (2012). "The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise," *Ear Hear.* **33**, 673–682.
- Culling, J., and Mansell, E. (2013). "Speech intelligibility among modulated and spatially distributed noise sources," *J. Acoust. Soc. Am.* **133**, 2254–2261.
- Culling, J. F. (2016). "Speech intelligibility in virtual restaurants," *J. Acoust. Soc. Am.* **140**, 2418–2426.
- Desjardins, J., and Doherty, K. (2014). "The effect of hearing aid noise reduction on listening effort in hearing-impaired adults," *Ear Hear.* **35**, 600–610.
- Farina, A. (2007). "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention* (Audio Engineering Society, New York), Vol. 122.
- Geurts, L., and Wouters, J. (2004). "Better place-coding of the fundamental frequency in cochlear implants," *J. Acoust. Soc. Am.* **115**, 844–852.
- Grange, J. A., and Culling, J. F. (2016). "The benefit of head orientation to speech intelligibility in noise," *J. Acoust. Soc. Am.* **139**, 703–712.
- Guillon, D., and Volle, M. (1987). "Gaze control in humans: Eye-head coordination during orienting movements to targets within and beyond the oculomotor range," *J. Neurophysiol.* **58**, 427–459.
- Hawley, M., Litovsky, R., and Culling, J. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.* **115**, 833–843.
- Hay-McCutcheon, M., Pisoni, D., and Kirk, K. (2005). "Audiovisual speech perception in elderly cochlear implant recipients," *Laryngoscope* **115**, 1887–1894.
- Hersbach, A., Arora, K., Mauger, S., and Dawson, P. (2012). "Combining directional microphone and single-channel noise reduction algorithms: A clinical evaluation in difficult listening conditions with cochlear implant users," *Ear Hear.* **33**, e13–e23.
- Hopkins, K., and Moore, B. (2011). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," *J. Acoust. Soc. Am.* **130**, 334–349.
- Jelfs, S., Culling, J., and Lavandier, M. (2011). "Revision and validation of a binaural model for speech intelligibility in noise," *Hear. Res.* **275**, 96–104.
- Kalikow, D., Stevens, K., and Elliott, L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* **61**, 1337–1351.
- King, A., Hopkins, K., and Plack, C. (2014). "The effects of age and hearing loss on interaural phase difference discrimination," *J. Acoust. Soc. Am.* **135**, 342–351.
- Koehnke, J., and Besing, J. (1996). "A procedure for testing speech intelligibility in a virtual listening environment," *Ear Hear.* **17**, 211–217.
- Kramer, S., Kapteyn, T., and Festen, J. (1998). "The self-reported handicapping effect of hearing disabilities," *Audiology* **37**, 302–312.
- Laske, R., Veraguth, D., Dillier, N., Binkert, A., Holzmann, D., and Huber, A. (2009). "Subjective and objective results after bilateral cochlear implantation in adults," *Otol. Neurotol.* **30**, 313–318.
- Laszig, R., Aschendorff, A., Stecker, M., Müller-Deile, J., Maune, S., Dillier, N., Weber, B., Hey, M., Begall, K., Lenarz, T., Battmer, R. D., Böhm, M., Steffens, T., Strutz, J., Linder, T., Probst, R., Allum, J., Westhofen, M., and Doering, W. (2004). "Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results," *Otol. Neurotol.* **25**, 958–968.
- Lavandier, M., and Culling, J. (2010). "Prediction of binaural speech intelligibility against noise in rooms," *J. Acoust. Soc. Am.* **127**, 387–399.
- Lavandier, M., Jelfs, S., Culling, J., Watkins, A., Raimond, A., and Makin, S. (2012). "Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources," *J. Acoust. Soc. Am.* **131**, 218–231.
- Litovsky, R., Parkinson, A., and Arcaroli, J. (2009). "Spatial hearing and speech intelligibility in bilateral cochlear implant users," *Ear Hear.* **30**, 419–431.
- Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (2006). "Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study," *Ear Hear.* **27**, 714–731.
- Loizou, P., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. (2009). "Speech recognition by bilateral cochlear implant users in a cocktail-party setting," *J. Acoust. Soc. Am.* **125**, 372–383.
- Loizou, P., and Kim, G. (2011). "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," *IEEE Trans. Audio Speech Lang. Process.* **19**, 47–56.
- MacLeod, A., and Summerfield, Q. (1987). "Quantifying the contribution of vision to speech perception in noise," *Br. J. Audiol.* **21**, 131–141.
- Mauger, S., Dawson, P., and Hersbach, A. (2012). "Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction," *J. Acoust. Soc. Am.* **131**, 327–336.
- Moore, B., and Glasberg, B. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Moore, B., Glasberg, B., Stoev, M., Füllgrabe, C., and Hopkins, K. (2012). "The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies (L)," *J. Acoust. Soc. Am.* **131**, 1003–1006.
- Müller, S., and Massarani, P. (2001). "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.* **49**, 443–471, available at <http://www.aes.org/e-lib/browse.cfm?elib=10189>.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- Peissig, J., and Kollmeier, B. (1997). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.* **101**, 1660–1670.
- Plomp, R. (1976). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," *Acta Acust. Acust.* **34**, 200–211, available at <http://www.ingentaconnect.com/content/dav/aaau/1976/00000034/00000004/art00004>.
- Plomp, R. (1986). "A signal-to-noise model for speech reception threshold of the hearing impaired," *J. Speech Hear. Res.* **29**, 146–154.
- Plomp, R., and Mimpen, A. (1979). "Improving the reliability of testing the speech reception threshold for sentences," *Int. J. Audiol.* **18**, 43–52.
- Ricketts, T., and Galster, J. (2008). "Head angle and elevation in classroom environments: Implications for amplification," *J. Speech Lang. Hear. Res.* **51**, 516–525.
- Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., and Barone, P. (2007). "Evidence that cochlear-implanted deaf patients are better multisensory integrators," *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7295–7300.
- Schorr, E., Fox, N., van Wassenhove, V., and Knudsen, E. (2005). "Auditory-visual fusion in speech perception in children with cochlear implants," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18748–18750.
- Schroeder, M. (1965). "New method of measuring reverberation time," *J. Acoust. Soc. Am.* **37**, 409–412.
- Strelnikov, K., Rouger, J., Barone, P., and Deguine, O. (2009). "Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants," *Scand. J. Psychol.* **50**, 437–444.
- Sumby, W., and Pollack, I. (1954). "Visual contribution to speech intelligibility in noise," *J. Acoust. Soc. Am.* **26**, 212–215.
- Summerfield, Q. (1987). "Some preliminaries to a comprehensive account of audio-visual speech perception," in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum, Hillsdale, NJ), Vol. 101, pp. 598–602.
- Summerfield, Q. (1992). "Lipreading and audio-visual speech perception," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **335**, 71–78.
- Van Hoesel, R. (2011). "Bilateral cochlear implants," in *Auditory Prostheses: New Horizons*, edited by F.-G. Zeng, A. N. Popper, and R. R. Fay (Springer, New York), pp. 13–57.
- Van Hoesel, R., Böhm, M., Pesch, J., Vandali, A., Battmer, R., and Lenarz, T. (2008). "Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing based strategies," *J. Acoust. Soc. Am.* **123**, 2249–2263.